UNIVERSIDAD: Universidad Nacional de La Plata (UNLP).

NUCLEO DISCIPLINARIO/COMITÉ ACADEMICO/OTROS TEMAS: Redes Académicas – Sistemas distribuidos y redes.

TITULO DEL TRABAJO: **DESARROLLO DE UN GRID INTERUNIVERSITARIO. APLICACIONES AL ESTUDIO DE PROBLEMAS MEDIOAMBIENTALES**

AUTOR(ES): Adrián Pousa, Ismael Rodríguez, José Petorutti, Gonzalo Ramírez Abella.

DIRECTORES: Ing. De Giusti Armando, Dr. Naiouf Marcelo.

CORREOS ELECTRÓNICOS DE LOS AUTORES: {apousa, ismael, josep, gramirez} @lidi.info.unlp.edu.ar.

PALABRAS CLAVES: Procesamiento distribuido y paralelo. Tecnología GRID. Simulación paramétrica

INTRODUCCIÓN

Es indiscutible la importancia actual del procesamiento paralelo y distribuido. En particular la utilización de arquitecturas de "cluster" de PCs es creciente por la relación costo/performance alcanzable. [BAK99].

A partir de los clusters aparecen nuevas arquitecturas paralelas, generadas por la interconexión de clusters a través de redes locales (LAN) o extendidas (WAN). Los problemas clásicos de performance, escalabilidad, heterogeneidad del hardware, balance dinámico de carga y *overhead* de comunicaciones, que caracterizan el estudio de algoritmos paralelos, reaparecen potenciados por las dificultades propias de la interconexión a través de una red no dedicada como puede ser Internet. [All01] [BAK99] [BOH02]

En esta línea se agrega en los últimos años la utilización de arquitecturas GRID donde sobre un sistema distribuido débilmente acoplado (por ejemplo en una red WAN sobre Internet) se comparten múltiples computadoras heterogéneas trabajando como una máquina paralela virtual. A diferencia de un multicluster dedicado, aquí las máquinas pueden compartir aplicaciones locales y remotas. [AHM03] [GOL02]

El cómputo paralelo en clusters se ha establecido desde hace varios años como una alternativa con grandes ventajas en cuanto a la relación costo/beneficio para resolver problemas numéricos y no numéricos con grandes requerimientos de rendimiento en el ámbito de procesamiento masivo de datos. [FOX99] La tendencia ha evolucionado de tal forma que las aplicaciones de cómputo intensivo aprovechan al máximo las ventajas de procesamiento creciente de las computadoras estándares de escritorio con costo aproximadamente constante y también han aprovechado la forma relativamente sencilla en la cual estas computadoras estándares pueden ser utilizadas para cómputo paralelo. [BAK99] [BAS99] [BOH02]

Por otro lado, también es posible la ampliación de las ideas de cómputo paralelo a más de un cluster, dando origen al cómputo paralelo intercluster. En este contexto, es más probable aún la redefinición de los principios de paralelización y/o el agregado de principios de paralelización para la obtención de rendimiento optimizado utilizando más de un cluster interconectado. [BOH02] [GOL02]

A partir de las arquitecturas de cluster y multicluster se generan nuevos modelos de cómputo ditribuido/paralelo que convergen en las tecnologías Grid. [MIN04]

El Grid ha surgido recientemente en el ámbito de la supercomputación para satisfacer las necesidades de ciertos proyectos científicos, bien por requerir una enorme capacidad de cálculo ("Computational Grids") o bien por manejar ingentes cantidades de datos ("Data Grids"). [All01] Pronto se ha visto que estos conceptos tienen un gran potencial también para determinadas aplicaciones comerciales y, actualmente, se prevé una posible

convergencia con las tecnologías del ámbito del comercio electrónico, con lo que se llegará a una implantación generalizada en todos los ámbitos ("Utility Grids"). [AHM03] [JOS03]

Durante este tiempo se ha producido un gran avance en el desarrollo de elementos software intermedios ("*middleware*"), necesarios en este tipo de aplicaciones, como ayuda adicional para la gestión de recursos distribuidos, seguridad, etc. Sin embargo, el desarrollo de aplicaciones que se asienten sobre la tecnología Grid sigue siendo, hoy por hoy, una labor casi artesanal, principalmente porque la distancia entre los conceptos requeridos por las aplicaciones y los proporcionados por el middleware es todavía demasiado grande. [BER03]

Entre las herramientas que actualmente son motivo de investigación y desarrollo en el área de Grid podemos mencionar:

- Ambientes y lenguajes específicos para el desarrollo de aplicaciones en entornos Grid.
- Algoritmos de monitoreo en tiempo real de los sistemas Grid, tanto a nivel de la utilización de los recursos del sistema como del comportamiento propio de las aplicaciones.
- Información y visualización del comportamiento dinámico del sistema Grid.
- Tolerancia a fallas en la arquitectura Grid, con transparencia para el usuario.

Hoy en día, la principal aplicación de la tecnología Grid es la supercomputación distribuida, es decir, poder utilizar recursos computacionales disponibles a través de Internet para la resolución de problemas con grandes requisitos de potencia de cálculo o de otros recursos como pueden ser capacidad de memoria, espacio de almacenamiento, etc. Esta tecnología se enfoca a resolver problemas tipo "gran desafío" (*grand challenge*) como pueden ser la simulación del clima a escala planetaria, análisis del genoma, simulación de procesos biológicos del ser humano, estudio de modelos de catástrofes naturales, análisis de datos epidemiológicos, estudios de astronomía, etc. [JOS03] [JUH04] [OGU03]

Sin embargo, y como se ha comentado anteriormente, Grid es un concepto mucho más amplio y revolucionario que va más allá de la simple realización de cálculos. De hecho, un área que emerge con gran fuerza en entornos en los que la diversidad de fuentes de información es grande, como es el caso de la biología en general, es el denominado "Data Grid", que viene a indicar el conjunto de tecnologías necesarias para integrar esa información y construir una red de recursos de procesamiento de la información realmente inter-operable. Aspectos claves de este gran área de "Data Grid" es la definición de Ontologías así como la necesidad de ejecutar "flujos de trabajos" (*workflows*) sobre diferentes datos para llegar a un resultado integrado. Claramente las aplicaciones de minería de datos (data *mining*) sobre grandes volúmenes de información lideran el interés por la explotación de la infraestructura de un Grid en esta área. [BER03] [JUH04] [TAL04]

En la actualidad, existen múltiples aproximaciones para la creación de una arquitectura Grid genérica mediante la definición de protocolos Grid estándares que permitan la interoperabilidad entre diferentes sistemas: definición de servicios, interfaces de aplicaciones y herramientas de desarrollo de software. Sin embargo, todavía no se ha impuesto ningún modelo concreto y aunque a nivel conceptual el acuerdo es grande, a nivel de aplicación el consenso es difícil. [CHE01] [FOS03] [OGU03]

La actividad en este campo a nivel mundial es muy importante y es en la fase de desarrollo de una nueva tecnología donde es vital la contribución de la comunidad científica. [HOS00] Precisamente este trabajo presenta una línea de desarrollo entre 4 Universidades Nacionales, que se integra con esfuerzos de otros países para configurar un Grid inter-universitario que permita compartir recursos de supercómputo a costos relativamente bajos.

Las aplicaciones que serán utilizadas como banco de pruebas de la tecnología Grid en este proyecto son:

- Algoritmos de simulación de incendios forestales.
- Algoritmos de estudio de modelos de inundaciones en ríos de llanura.

En una etapa posterior se agregarán:

- Algoritmos de reconstrucción 3D (en particular aplicables en Medicina).
- Algoritmos de reconocimiento de secuencias (en particular de ADN o genómicas).

DESARROLLO

Estudios realizados en las capas iniciales del Software de GRID

En el marco del proyecto CyTED, se ha realizado un análisis comparativo de diferentes *middlewares* para la construcción y puesta en marcha de sistemas GRID. En particular, se han realizado comparaciones y análisis entre Globus Toolkit y GLite.

Analizando GLite, se llegó a la conclusión de que este *middleware* posee altos requerimientos de hardware y/o software; debido a su arquitectura distribuida en diversos equipos (característica no siempre soportada en algunos casos del ámbito académico). Por otra parte, GLite es un proyecto centralizado (liderado por el CERN) que torna dificultoso que los grupos realicen aportes. Por estos motivos, se ha tomado la decisión de (en un principio) realizar la instalación de GRID utilizando el *middleware* Globus Toolkit 4.0.

Análisis de las posibilidades de Globus 4 y su instalación

Globus Toolkit 4 es un *framework* de componentes *open source* para la creación, puesta en marcha y administración de GRID y aplicaciones sobre la misma.

Se basa en diversos standards en evolución, tales como Web Services Resource Framework (WSRF), Open Grid Service Infrastruscture (OGSI), Grid Security Infrastructure (GSI), Simple Object Access Protocol (SOAP), Extensible Markup Lenguaje (XML), y otros protocolos de base. [VLA05]

Globus Toolkit 4 ofrece muchos servicios, en su mayoría basados en web services, y otros como *daemons* o aplicaciones escritos en su mayoría en lenguaje C o J2EE.

Algunos de los servicios de GT4 que no están basados en Web Services se mantienen por razones de compatibilidad y migración con las versiones anteriores del toolkit. Una comparación en relación a estas características entre versiones anteriores de Globus Toolkit es la siguiente:

- Globus Toolkit 2 no posee servicios basados en web services.
- Globus Toolkit 3 integraba web services y no-web services. Esta versión está basada en el Standard OGSI.
- Globus Toolkit 4 está íntegramente basado en servicios web, salvo excepciones de componentes del toolkit que no es posible implementarlas con Web Services. Esta versión se adapta al estándar WSRF.

Un aspecto importante en los sistemas GRID es el de seguridad. La seguridad de GRID se basa en la infraestructura de clave publica (PKI), utilizando Entidades Certificantes para la autorización y autenticación de los nodos y usuarios. Globut Toolkit 4 posee en su distribución un sistema de entidad certificante llamada SimpleCA, que no es recomendado para sistemas en producción. [VLA05]

Instalación de GRID basada en el middleware Globus Toolkit 4

La ultima versión de Globus Toolkit 4 se puede obtener del sitio oficial de Globus (http://www.globus.org/). En el mismo se encuentran binarios y paquetes de instalación para diferentes sistemas operativos. Además de los paquetes binarios se puede obtener código fuente y compilarlo para aquellos sistemas operativos para los cuales no se encuentra disponible una distribución binaria.

Los requerimientos de software para instalación, entre otros, son:

- J2SE 1.5 o superior
- Apache Ant
- Motor RDBMS para su posterior utilización por el servicio Reliable File Transfer (RFT)
- Sincronización del reloj de los nodos, usando un protocolo de sincronización como NTP
- Dependiendo de la distribución del Sistema Operativo, se debe instalar el servicio Inetd (Xinetd). El mismo es utilizado por los servicios No-WS del toolkit.

La planificación, instalación y configuración de una GRID en general no es una tarea sencilla. Se debe planificar la topología, instalar el middleware, obtener y firmar certificados de seguridad para nodos de la GRID y usuarios ante la entidad certificante (la cual también puede residir dentro de nuestra topología) y la configuración de servicios.
Se adaptó la topología multicluster existente para soportar una infraestrucutra GRID, configurando nodos GRID como cabeza de cluster. Para la implementación de los nodos GRID, se optó utilizar Globus Toolkit 4 sobre plataforma GNU/LINUX en su distribución Fedora Core 5. La instalación del middleware consistió en la configuración de todos sus componentes: GridFTP, RFT, GRAM, WS-GRAM, RLS, MDS.
Se montó una entidad certificante (CA) utilizando OpenCA, físicamente ubicada en la ciudad de La Plata.
Tanto los nodos GRID como los nodos de los clusters, fueron configurados con un Resource Manager “Torque” y un Cluster Scheduler “Maui”. Los nodos del cluster están configurados para trabajar con las librerías de MPI. Las herramientas anteriormente mencionadas administran los recursos y solicitudes de procesamientos sobre los clusters.
Para integrar la infraestructura multicluster al GRID, se configuró WS-GRAM interactuando con Torque logrando así el acceso a los recursos de los cluster a través del GRID.
Como herramienta de monitero se utilizó Ganglia.
Se realizaron pruebas de los componentes brindados por el middleware GRID, teniendo en cuenta la comunicación multicluster, obteniendo los resultados favorables.

Las tareas que se realizaron durante la instalación y configuración del middleware fueron:

- Instalar el sistema operativo
- Instalar J2SE 1.5
- Realizar la sincronización por NTP
- Instalar el toolkit GT4
- Implementar la configuración de la CA en dos de los nodos
- Configurar los certificados de nodo
- Realizar la creación de usuarios y configuración de certificados de usuarios de grid
- Mapear los usuarios de grid a usuarios de sistema
- Realizar la configuración de WS-GRAM
- Configurar GridFTP
- Instalacion y configuración de un Resource Manager “Torque” y un Cluster Scheduler “Maui”.

La configuración global considerando la conexión con otras universidades es la siguiente:

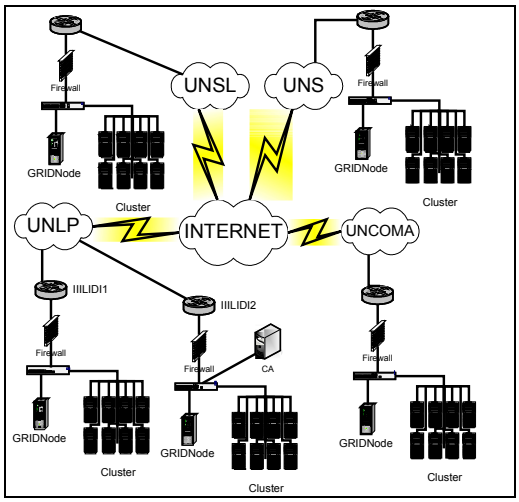

Figura1: Topologia

Es importante mencionar que se encontró un inconveniente relacionado con las direcciones IP privadas (no ruteables). Uno de los servicios que mostró este problema fue GridFTP: al realizar una transferencia, en un momento la dirección privada es encapsulada dentro del paquete de datos; al llegar el paquete al router, éste no puede realizar la conversión por NAT pasando la información de direccion IP privada a un nodo que al querer responder no la puede rutear. Para solucionar este problema, se decidió utilizar direcciones IP públicas, instalando las GRID en diferentes redes. Con el fin de facilitar la tarea de instalación, se desarrollaron una serie de *scripts* parametrizados que automatizan el proceso.

Pruebas de conexión y servicios con otras Universidades

Se realizaron pruebas de conexión con tres universidades:

- Universidad Autónoma de Barcelona (España)
- Universidad Nacional del Sur (Bahía Blanca)
- Universidad Nacional de Comahue (Neuquén)

En los tres casos se realizaron con éxito pruebas sobre los mismos servicios probados en la red configurada por nosotros, quedando en evidencia el problema de las direcciones IP privadas en aquellas universidades que cuentan con ese tipo de arquitectura.

Un tema importante a tener en cuenta es el cruce de certificados para autenticación. Esto se da cuando interactúan varios nodos de grid que no necesariamente tienen una misma entidad certificante; al intercambiar certificados entre dos nodos queda establecida una relación de confianza entre dos entidades certificantes diferentes.

Pruebas de ejecución utilizando MPI sobre nodos de GRID

Existe una distribución de la librería MPI adaptada para utilizar en GRID (MPICH-G2, que puede descargarse del sitio http://www-unix.mcs.anl.gov/mpi/mpich/). Se encuentran dos versiones de MPICH (MPICH1 y MPICH2), y la versión para utilizar en Grids es MPICH1 para Globus llamada MPICH-G2.

MPICH-G2 está preparada para usarse en Globus Toolkit 2 bajo sevicios no basados en web services, pero de todos modos puede utilizarse en Globus Toolkit 4 ya que esta última versión mantiene estos servicios por razones de compatibilidad con las versiones anteriores.

Para utilizar MPICH-G2 en Globus Toolkit 4 se debe configurar el Globus Gatekeeper, que es parte de los servicios no basados en web services de Globus Toolkit 2 y es el encargado de autenticar una tarea y su propietario para luego trasladar el requerimiento al JobManager encargado de procesarlo.

En la práctica se observó que máquinas con MPICH y Globus Toolkit instalado no pueden utilizarse con aquellas que tengan MPICH y no tengan instalado Globus Toolkit, este inconveniente se debe a que la distribución de MPICH-G2 compila utilizando librerías de Globus además de utilizar un puerto en particular para recibir pedidos de procesamiento.

Se realizó la instalación de MPICH-G2 en los diferentes nodos compilado para Globus Toolkit. A diferencia de mpich estándar, en el caso de mpich para Globus se debe generar un Proxy para poder ejecutar un programa, al cual Globus le delega la responsabilidad de autenticación (la generación del Proxy es necesaria para utilizar cualquier servicio de Globus). Otra diferencia con la ejecución con mpich estandar es que para la ejecución del programa se debe generar un archivo RSL que indica la distribución de la tarea entre los distintos nodos; esto significa que si el programa necesita cinco procesos y se cuenta con dos nodos, se indica en el archivo RSL que un nodo ejecutará dos de estos procesos y el otro nodo los tres restantes. Finalmente, se necesita un comando de Globus para correr el programa utilizando el archivo RSL.

Se ejecutaron diversos programas clásicos de procesamiento distribuido y paralelo utilizando nodos locales y remotos sobre el grid. Las pruebas realizadas fueron satisfactorias e intentaron probar el funcionamiento de la distribución de MPICH-G2, así como servir de base para la ejecución de aplicaciones más complejas.

CONCLUSIONES / PROYECTO EN DESARROLLO

- El desarrollo de una infraestructura Grid que permita interconectar y utilizar coordinadamente recursos de varias Universidades del país es claramente de gran utilidad. Para esto se ha avanzado en las pruebas de conexión de redes de las Universidades del Sur, Comahue, y La Plata, y próximamente con San Luis.

- Las aplicaciones elegidas inicialmente (aunque no son excluyentes) para probar esta infraestructura, son de importancia para el país ya que los problemas medioambientales (incendios e inundaciones) que se han elegido tienen impacto social y económico. Su estudio sobre una arquitectura de procesamiento importante permitirá obtener predicciones e información útil para la toma de decisiones (incluso eventualmente en tiempo real).
- En general el desarrollo del software de soporte de una infraestructura distribuida tal como la propuesta es complejo, no tiene soluciones "únicas" o suficientemente probadas y requiere una intensa tarea experimental.
- A su vez la construcción de aplicaciones (algoritmos paralelos) que exploten eficientemente la potencia de cálculo de un Grid requiere un gran esfuerzo ya que se trata de un tipo de Sistema de procesamiento nuevo, con heterogeneidad en los procesadores y en las comunicaciones y sin un sistema operativo único.
- Las tareas que continúan este trabajo son importantes y se centran en el despliegue de las máquinas en Argentina, la incorporación de otras Universidades y la vinculación de los nodos con esquemas similares en 14 países de Iberoamérica que están incluidos en el proyecto CyTED "Tecnología GRID como motor de desarrollo regional" actualmente en desarrollo.

REFERENCIAS

[AHM03] Ahmar Abbas. "Grid Computing : Practical Guide To Technology & Applications". Programming Series. Charles River Media; 1 edition (December, 2003)

[All01] Allcock W., Bester J., Bresnahan J., Chervenak A., Foster I., Kesselman C., Meder S., Nefedova V., Quesnel D., Tuecke S. "Secure, efficient Data Transport and Replica Management for high-performance data-intensive computing". 18th IEEE Symposium on Mass Storage Systems. 2001.

[BAK99] Baker M., R. Buyya. "Cluster Computing at a Glance". R. Buyya Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA,pp.3-47,1999.

[BAS99] Basney J., M. Livny. "Deploying a High Throughput Computing Cluster". R. Buyya Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA, pp. 116-134, 1999.

[BER03] Berman F.(Editor), Fox G.(Editor), Hey A.(Editor). "Grid Computing: Making The Global Infrastructure a Reality". John Wiley & Sons (April 8, 2003).

[BOH02] Bohn C, Lamont G. “Load Balancing for Heterogeneous Clusters of PCs”. Future Generation Computer Systems, Elsevier Science B.V., Vol 18, 2002, pp 389-400.

[CHE01] Chervenak A., Foster I., Kesselman C., Salisbury C., Tuecke S. “The data Grid: Towards and Architecture for the Distributed Mannagement and Analysis of Large Scientific data Sets”. Journal of Network and Computer Applications, 2001. 187-200.

[FOS03] Foster I., Kesselman C., Kaufmann M. “The Grid 2: Blueprint for a New Computing Infrastructure”. The Morgan Kaufmann Series in Computer Architecture and Design. 2 edition (November 18, 2003).

[FOX99] Baker M., Fox G. "Metacomputing: Harnessing Informal Supercomputers". R. Buyya Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA, pp. 154-185, 1999.

[GOL02] Goldman. “Scalable Algorithms for Complete Exchange on Multi-Cluster Networks”. CCGRID'02, IEEE/ACM, Berlin, p. 286 - 287, 2002.

[HOS00] Hoschek W., Jaen-Martinez J., Samar A., Stockinger H., Stockinger K. “Data Management in an Internatioonal Data Grid Project”. International Workshop on Grid Computing, Springer-Verlag 2000.

[JOS03] Joseph J., Fellenstein C. ”Grid Computing”. On Demand Series. IBM Press (December 30, 2003).

[JUH04] Juhasz Z. (Editor), Kacsuk P. (Editor), Kranzlmuller D. (Editor). “Distributed and Parallel Systems: Cluster and Grid Computing”. The International Series in Engineering and Computer Science. Springer; 1 edition (September 21, 2004)

[MIN04] Minoli D. “A Networking Approach to Grid Computing”. Wiley-Interscience (October 15, 2004).

[OGU03] Ogura S, Nakada H, Matsuoka S. “Evaluation of the inter-cluster data transfer on Grid environment”. Proceedings of CCGrid 2003 , pp. 374-381, May 2003.

[TAL04] Talwar V., Agarwalla B.,Basu S., Kumar R., Nahrstedt K. “Architecture for Resource Allocation Services supporting Interactive Remote Desktop Sessions in Utility Grids”. Proceedings of the 2nd workshop on Middleware for grid computing,2004.

[VLA05] Vladimir Silva. “Grid Computing For Developers”. Programming Series. Charles River Media; 1 edition (December 15, 2005).

Grid.Org: http://www.grid.org/

Grid Computing Info Centre (GRID Infoware): http://www.gridcomputing.com/

IEEE Task Force on Cluster Computing http://www.ieeetfcc.org/

Colección “IEEE Transaction on Parallel and Distributed Systems”.